



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/884,403	06/19/2001	Shamim A. Alpha	27252.2	2286

7590

08/03/2005

Petar Kraguljac
McDonald Hopkins Co LPA
600 Superior Avenue E
Suite 2100
Cleveland, OH 44114

EXAMINER

SPOONER, LAMONT M

ART UNIT

PAPER NUMBER

2654

DATE MAILED: 08/03/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/884,403

Applicant(s)

ALPHA, SHAMIM A.

Examiner

Lamont M. Spooner

Art Unit

2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 19 June 2001.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-20 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-20 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 19 June 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: _____

DETAILED ACTION

Specification

1. The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed.

Claim Objections

2. Claim 10 is objected to because of the following informalities:

In claim 10, line 2, "the probability data" has antecedent issues.

Appropriate correction is required

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. Claims 1-9, 11-13, 15, 16, and 20 are rejected under 35 U.S.C. 102(b) as being anticipated by Pon et al. (Pon, US 6,047,251).

As per **claims 1 and 2**, Pon teaches a system for automatically determining a language of a document from a set of candidate of languages, the system comprising:

logic (C.7.lines 33-35-his subroutine) for setting a negative assumption value (C.7.lines 36, 37-his setting of an initial "confidence statistic") indicating the document is not one of the candidate languages (ibid, C.7.lines 1, 2, C.7.lines 35-37,-interpreted that at a "zero" confidence level, the document is deemed not one of the candidate languages);

an extractor for extracting a character string from the document (C.7.lines 38-40);
and

a language analyzer (Fig. 4 item 106, Fig 5) for determining a probability value that the character string does not belong to the candidate languages (C.6.line 65-C.7.line 22-the "statistic that indicates whether a selected word is in a chosen language", wherein the "probability that a character string belongs to each of the candidate languages result inherently determines the value that a character string does not belong, claim 2) and includes logic for adjusting the negative assumption value based on the probability value (C.7.lines 39-41-his "updating"), the language analyzer determining that the document is one language of the candidate languages when the negative assumption value passes a threshold value (C.8.lines 1-4, his "region" as the document, his current subzone for the region "is likely to be the language of the region, C.8.lines 5-25-use of the threshold, C.9.lines 10-12-entire document).

As per **claim 3**, Pon teaches claim 2, and further teaches further including logic for retrieving the probability value from the probability data that corresponds to the character string (Fig. 5 item 130 and return logic).

As per **claim 4**, Pon teaches claim 1, and further teaches further including an information retrieval engine for retrieving documents in response to a search request, the documents retrieved being analyzed by the language analyzer (C.4.lines 29-32-downloading inherent to a search request, Fig. 2 item 52-identify language).

As per **claim 5**, Pon teaches claim 1, and further teaches wherein the logic for adjusting includes logic for combining the negative assumption value (C.7.lines 37, 38-

Art Unit: 2654

the "initial value") with the probability value (C.7.lines 39-45-his "region confidence factor" as a statistical value is added to the "initial" value).

As per **claim 6**, Pon teaches claim 1, and further teaches wherein the language analyzer further includes iteration logic for causing the extractor to extract another character string from the document if the negative assumption value fails to pass the threshold value (Fig 7- from the "select region r" item 158 to the "more regions" the loop is interpreted as iteration logic, C.8.lines 12-25-the appending of another subzone, which includes another character string from the document appended, C.7.lines 50-55-"regions" and "zones" and words therein).

As per **claims 7, 8 and 13**, Pon teaches a method of determining a language of a document from a set of candidate languages, the method comprising the steps of:

setting a null hypothesis to a true value for each candidate language indicating the document is not in the candidate language and setting a false value (C.7.lines 36, 37-his setting of an initial "confidence statistic", C.7.lines 1, 2-his 1, as true, and 0, as false, value, claim 13);

extracting a text string from the document (C.7.lines 38-40);

determining a contrary probability for each candidate language that the text string does not belong to the candidate language (C.6.line 65-C.7.line 22-the "statistic that indicates whether a selected word is in a chosen language", wherein the "probability that a character string belongs to each of the candidate languages result inherently determines the value that a character string does not belong);

adjusting the null hypothesis for each candidate language with the contrary probability corresponding to the candidate language (C.7.lines 39-45-his "updating"- "value stored in the accumulator"-as the null hypothesis); and

determining the document is one language from the candidate languages when the null hypothesis for the one language is disproved by approaching the false value (C.7.lines 40-45-disproval interpreted the accumulation away from the true value above, C.8.lines 1-4, his "region" as the document, his current subzone for the region "is likely to be the language of the region, C.8.lines 5-25-use of the threshold, C.9.lines 10-12-entire document, wherein the accumulation).

As per **claim 9**, Pon teaches claim 8 and further teaches repeating the extracting step for a different text string from the document and repeating the method until the null hypothesis is disproved for one of the candidate languages by passing the threshold value ((Fig 7- from the "select region r" item 158 to the "more regions" the loop is interpreted as iteration logic, C.8.lines 12-25-the appending of another subzone, which includes another character string from the document appended, C.7.lines 50-55-"regions" and "zones" and words therein).

As per **claim 11**, Pon teaches claim 7 and further teaches identifying the document based on a search request (C.4.lines 29-32-downloading inherent to a search request, Fig. 2-subsequent processing).

As per **claim 12**, Pon teaches claim 7 and further teaches extracting a plurality of sequential characters that form the text string (C.4.lines 64-67, C.7.lines 6-7-Examiner interprets word to comprise sequential characters).

As per **claims 15 and 16**, claims 15 and 16 set forth limitations similar to claims 1 and 7, and therefore are rejected for the same reasons and under the same rationale.

As per **claim 20**, claim 20 sets forth limitations similar to claims 4 and 11, and therefore is rejected for the same reasons and under the same rationale.

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. Claims 10, 14 and 17-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pon in view of Elworthy (US 6,125,362).

As per **claim 10**, Pon teaches claim 7 but lacks further teaching pregenerating probability data corresponding to each candidate language, the probability data including a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages.

However, Elworthy teaches pregenerating probability data corresponding to each candidate language (C.2.lines 30-35-his "classification" as the candidate language), the probability data including a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages (ibid, his "determined probability that an element or group of elements belongs to a classification" is interpreted as occurrence probability, C.2.lines 30-38, the comparison with probability values interpreted as the normalization). Therefore, at the time of the invention, it would

Art Unit: 2654

have been obvious to modify Pon with Elworthy by using pregenerating data as a probability step. The motivation for doing so would have been to develop a increasing accurate method in classifying data (C.2.lines 16-20).

Claim 17 sets forth limitations similar to claim 10, and is thus rejected for the same reasons, and under the same rationale, wherein Elworthy further teaches contrary probability of a character string in one language is determined based on an occurrence frequency of the character string in the one language influenced by a total occurrence frequency of the character string in all the candidate languages (C.8.lines 27-31-his "tokens" as character strings, Fig. 14a, b, c, C.13.lines 43-58-wherein the "probability" values inherently contain contrary probability values).

As per **claim 18**, Pon and Elworthy make obvious claim 17, Elworthy further teaches determining the occurrence frequency of each character string based on a sample set of documents provided for each of the candidate languages (C.7.line 65-C.8.line 7).

As per **claim 19**, Pon and Elworthy make obvious claim 17, Elworthy further teaches wherein the contrary probability of the character string in one language is normalized by the total occurrence frequency of the character string in all the candidate languages (C.8.lines 27-31, C10.line 15-C.11.line 37, especially C.10.lines 50-57-his "frequency of all word tokens in M, and $p(m)$ as the normalization).

As per **claim 14**, Claim 14 sets forth limitations similar to claims 17, 18, and 19, and is thus rejected for the same reasons, and under the same rationale.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

- van den Akker (US 6,415,250) teaches automatically identifying a language using predetermined portions of words and probabilities methods.
- Martino et al. (US 6,002,998) teaches determining languages from text by probabilistic word tables, and word comparison.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Lamont M. Spooner whose telephone number is 571/272-7613. The examiner can normally be reached on 8:00 AM - 5:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571/272-7602. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Donald L. Stearn
PATENT EXAMINER
AU 2654

Application/Control Number: 09/884,403
Art Unit: 2654

Page 9

lms
7/26/05